DOCUMENT RESUME

ED 045 202 24 PS 004 019

AUTHOR	Bayder, Nicholas E.: And Others
TITLE	Preliminary Analysis on Kindergarten and First Grade Follow Through Test Results for 1968-69.
INSTITUTION	Far West Lab. for Educational Research and Development, Berkeley, Calif.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C. Division of Educational Laboratories.
REPORT NO	OCCAS-RR-2
BUREAU NO	BR-0-0481
PUB DATE	Feb 70
CONTRACT	OEC-0-70-3362
NOTE	26p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.40
DESCRIPTORS	Cognitive Processes, Community Study, Comparative Analysis, Concept Formation, *Grade 1, *Intellectual Development, *Kindergarten, Tables (Data), Testing, *Test Reliability, *Test Results, Tests
IDENTIFIERS	Project Follow Through, Responsive Follow Through Program

ABSTRACT

This paper is a preliminary report of objective test results made by approximately 300 poor and 100 non-poor kindergarten children and by 417 first grade children during 1968-1969. Tests administered to the kindergarten classes in the fall of 1968 and again in the spring of 1969 were: short-form Wechsler Preschool and Primary Scale of Intelligence (WPPSI), Caldwell Preschool Inventory (PSI), Innovative Behavior Test (IB), and Categories Test (C). The Metropolitan Readiness Test was administered in pre-post fashion to the first grade classes. The test results were used to measure changes in a child's intellectual ability as a result of one year's schooling in a Responsive Follow Through classroom. Except for four instances, average tests scores made by groups of both poor and non-poor kindergarten Follow Through children increased for all tests in every community. That is, over the seven to nine month time period, kindergarten children in all communities increased in intelligence, the ability to form concepts, the ability to solve problems and the ability to categorize. Results for first grade children were also encouraging. There were considerable increases in Metropolitan test scores for both first grade communities from the beginning to the end of the school year. (Author/NH)

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

PA-24
BR-0-0481
OEC-0-70-336

ED045202

PRELIMINARY ANALYSIS ON KINDERGARTEN AND FIRST GRADE FOLLOW THROUGH TEST RESULTS FOR 1968-69

OCCASIONAL RESEARCH REPORT NUMBER 2

FEBRUARY 1970

Compiled and Written by the

Evaluation Staff

Education Beginning at Age Three

Far West Laboratory for Educational Research and Development

PS004019

FOR LIMITED DISTRIBUTION

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Glen Nimnicht

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

Far West Laboratory for Educational Research and Development
Level II
Date: February 1970
Author: N. F. Rayder

TABLE OF CONTENTS

		Page
I.	Introduction	1
II.	The Testing Program	2
	A. Kindergarten	2
	B. First Grade	6
III.	Reliability of the Tests	7
IV.	Kindergarten Results by Community	9
	Community A	9
	Community B	11
	Community D	12
	Community E	13
	Community F	14
	Community G	15
	Community H	16
	Summary of Kindergarten Test Results	17
V.	First Grade Results by Community	21
	Community I	21
	Community J	21
VI.	Overview	22

The Laboratory was established through a Joint Powers Agreement in February 1966. Signatories as of June 1969 include:

The Regents of the University of California
The California State Board of Education
The Trustees of the California State Colleges
The County Superintendent of Schools of the County of Monterey
The Board of Education of the San Francisco Unified School District
The Regents of the University of Nevada
The Nevada State Board of Education
The Board of Regents of the University of Utah
The Utah State Board of Education

PRELIMINARY ANALYSIS ON KINDERGARTEN AND FIRST GRADE
FOLLOW THROUGH TEST RESULTS FOR 1968-69
OCCASIONAL RESEARCH REPORT NUMBER 2

FEBRUARY, 1970

I. Introduction

During 1968-69 ten communities throughout the United States cooperated with the Far West Laboratory for Educational Research and Development to test a model Follow Through Program in kindergarten and first grade. The model, called the Responsive Follow Through Program, has two major objectives.

1. To help children develop a positive self-image, and
2. To help children develop their intellectual abilities.

This paper addresses itself to the second objective and describes the testing program and instruments used to measure changes in intellectual ability of eight kindergarten and two first grade Follow Through classes during the 1968-69 school year. First, the testing program is described. Test results for each community are then presented and discussed. This is followed by a general summary and discussion. Finally, limitations of the testing procedures and plans for further analysis of the data are presented.

Before describing the methods used to measure intellectual development in kindergarten and first grade children, efforts to evaluate the Responsive Follow Through model should be put into perspective.

Educators and researchers tend to measure the effectiveness of an educational system or model by pointing to gains in objective test scores. Emphasis is placed on changes made on objective tests for at least two reasons:

1. The objective testing movement is so ingrained in our society that it is natural for an educator, researcher or parent to look solely at objective test results as indicators of a program's success, and
2. Because of the emphasis on objective testing of cognitive skills,

development of ways to measure other educational outcomes have received only minor attention. For example, there is no "widely accepted" instrument to measure a child's self-concept, his desire for learning, or his attitude toward school. Until instruments are developed to measure these "affective" goals, we will continue to rely on objective test results for evaluation.

One other point should be made: The word "Preliminary" is used in the title of this report for a special reason. Because of the long-term nature of our goals, a one-year assessment of the degree to which the Responsive Model achieves these goals cannot be adequate. For this reason, various Follow Through model developers, called sponsors, and the National Follow Through office have placed a moratorium on between-sponsor, experimental-control group comparisons. Comparisons made between sponsors, other than describing and contrasting methods, are premature at this time.

II. The Testing Program

To measure the development of a kindergarten child's intellectual ability during his nine month experience in the Responsive Model, a four-test battery reflecting some of the program's objectives was administered to the eight Follow Through kindergarten classes in the fall of 1968 and again in late Spring of 1969.

A separate instrument appropriate for first grade, was administered in pre-post fashion to first grade Follow Through classes. Comparison classes, selected for similarity to the Follow Through classes, were also pre-and post-tested with the same battery. For reasons already discussed, the test results of comparison groups are not presented in this paper.

A. The kindergarten test battery was composed of four instruments;

1. A short form of the Wechsler Preschool and Primary Scale of

Intelligence (WPPSI).[1] The WPPSI was designed to measure intelligence in young children aged four to six and a half. Subtests are individually administred by a trained person. Only four subtests were selected for administration because they were the best measures of the objectives of the model and of the subtests in the WPPSI. They provide the most reliable estimate of intelligence. The four subtests that made up the short-form WPPSI were:

	WPPSI Subtest	Related Responsive Objective
a.	Vocabulary	Language development
b.	Similarities	Concept formation
c.	Picture Completion	Perceptual acuity
d.	Block Design	Problem solving

The Vocabulary subtest requires the child to answer such questions as "What is a shoe?" and "What does "chisel" mean?" This subtest measures word knowledge and potential for dealing with symbols.

The Similarities subtest measures verbal concept formation by requiring the child to make verbal associations within such categories as clothing, foods and toys. The Vocabulary and Similarities subtests combine to yield a Total Verbal Score.

The Picture Completion subtest is a measure of verbal concentration. It requires a child to find missing pieces in a picture and to be aware of objects in his world.

The Block Design subtest requires the child to analyze and reproduce an abstract design. It also appears to measure insight into space relations. The Picture Completion and Block Design subtests can be combined to give a score on Total Performance.

---

1. Wechsler, David, "Manual for the Wechsler Preschool and Primary Scale of Intelligence", The Psych. Corp., N. Y., N. Y. 10017

The combination of the Total Verbal and Total Performance scores was used as a measure of a child's intelligence and is referred to as the Total WPPSI score. To get the Total WPPSI score, raw scores for each of the four subtests were converted to scale scores using different age tables. These four scale scores were then added. Based on the standardization sample, the general population of children would have an average Total WPPSI score of 40. The standard deviation of these scores would be around 10.

The WPPSI is a relatively new instrument and research on its accuracy is scant. However, because construction and administration procedures parallel procedures used in developing other Wechsler scales, the WPPSI enjoys the Wechsler reputation. Wechsler scales are reliable and valid for measuring intelligence and predicting scholastic success.

2. The Caldwell Preschool Inventory (PSI)[2]. The PSI was developed to be used with children from three to six years of age to measure achievement in areas regarded essential for success in school. The author, Bettye M. Caldwell, claims that the test is not culture-free, but was designed to permit educators to highlight the degree of a child's intellectual deficit. The 85-item instrument yields four major factors plus a total score. The factors accounting for most variability are called "Concept-Activation." The "concepts" are composed of ordinal or numerical relations and sensory attributes such as form, color, size, shape and motion. The "activation" involves "being able to call on established concepts to describe or compare attributes"; for example, to relate shapes to objects or to "execute motorically

---

[2]Caldwell, Bettye M., "The Preschool Inventory, Technical Report", Cooperative Test Division - Educational Testing Service, Princeton, N. J., 1967.

some kind of spatial concept" such as reproduction of geometric designs or drawing the human figure. The other two factors are Personal Social Responsiveness--involving knowledge about the child's own personal world--and Associative Vocabulary--requiring the demonstration of awareness of the connotation of a word by carrying out some action.

Raw Scores are converted to percentiles using norming tables provided in the manual. Test reliability reported in the PSI manual is scant. A .98 correlation was obtained between the score on the complete version and score on the shortened version for the original standardization sample of 171 children. A reliability correlation of .95 (a split-half corrected using the Spearman-Brown formula) was obtained for the shortened version using scores made by the same standardization sample.

3. The Innovative Behavior Test (IB). This instrument is one subtest in the Cincinnati Autonomy Test Battery developed by Thomas J. Banta[3]. It was designed to measure problem-solving ability in young children by having them trace different ways to get a dog to a bone. The test battery has not been published; consequently, there is no reliability information available. Test results are reported in raw score form and range from 0 to 30.

4. The Categories Test ("C" test). This test was internally developed to test the ability of young children to make simple classifications. For example, the child is handed a flashlight and shown nine objects. One of the objects is a light bulb that "goes with" the flashlight. The child is correct if he points to or names the correct response. Results are also in raw score form and range from 0 to 9.

PS004019

[3]Banta, Thomas J., "Tests for the Evaluation of Early Childhood Education: The Cincinnati Autonomy Test Battery (CATB)", (in published paper), University of Cincinnati, Cincinnati, Ohio 1968.

Although the Categories test has been used in other studies, the reliability has not been reported[4].

B. First Grade Tests

The test administered to the two first grade classes was the Metropolitan Readiness Test[5].

The Metropolitan Readiness Tests were developed to be administered at the end of kindergarten or the beginning of first grade to provide an objective reliable measure of readiness to do first grade work. The test is group administered and a child's total score is made up of items in six areas:

1. Word meaning - range of Vocabulary and verbal concepts,
2. Listening - comprehension of phrases and sentences, sustained attention, capacity for inference,
3. Matching - visual perception of similarities and differences in word forms and figures,
4. Alphabet - ability to recognize lower-class letters of the alphabet by name,
5. Numbers - number maturity, vocabulary, counting, recognition of written numerals, interpreting number symbols, simple numerical problems, and
6. Copying - combination of visual perception and motor control that is important in learning to write.

For interpreting test results the test manual provides five levels of "readiness status" as well as percentile ranks for beginning-first grade and end-of-kindergarten groups. The norms are based on a national sample

---

[4]Nimnicht, Glen P. and others, "Interim Report: Research on the New Nursery School", Colorado State College, Greeley, Colorado, Dec. 1967.

[5]Hildreth, G. H., Griffins, N.L. and McGauvran, "Metropolitan Readiness Tests - "Manual oif Directions", Harcort, Brace and World, Inc., Test Department, N. Y

of 12,231 children. In spite of their efforts to develop representative norms the authors caution that "the final standarization group may be slightly superior to the national average with respect to median income and average schooling of adults in the communities..."

III. Reliability of the Tests

Before test results can be discussed, it is important to know how "reliable" the test is. Reliability is how consistently a test measures whatever it does measure. One way to obtain evidence of test reliability is to calculate a correlation coefficient between successive administrations of the test.

To some extent the magnitude of the reliability coefficient arrived at by correlating successive administrations will depend on the length of time between the administrations. If the re-administrations are very close together in time, the results may be influenced by practice. When the time between successive administrations is long, part of the fluctuation in scores may be due to different rates of growth or change in whatever the test measures.

To get evidence of the reliability of the WPPSI, IB and C tests used in kindergarten classes, correlation coefficients were calculated on successive administrations of the tests a month apart. Since PSI scores are transformed to percentile scores before the correlations were run, the correlation between successive administrators of the PSI is not reported.

The correlations calculated on successive administrations one month apart for the WPPSI, IB and C tests are reported in Table I.

TABLE I

CORRELATION COEFFICIENTS CALCULATED BETWEEN SUCCESSIVE ADMINISTRATIONS OF THE WPPSI, INNOVATIVE BEHAVIOR AND CATEGORIES TEST FOR POOR AND NON-POOR CHILDREN

TEST	POOR*		NON-POOR	
	N	r	N	r
WPPSI	(142)	.78	(72)	.77
IB	(121)	.34	(48)	.40
C	(111)	.28	(31)	-.15

As shown in Table I the WPPSI has respectable reliability when used with poor and non-poor children. The Innovative Behavior and Categories tests, however, demonstrate extremely low reliability for both poor and non-poor children. This reflects the developmental stage these tests are in along with the limited range of total possible scores a child can get on each of them.

Based on the evidence reported above and on published reliability data on the PSI, the WPPSI and PSI are reliable instruments when used with poor and non-poor children. As indicated by the extremely low reliability coefficients on successive administrations, the Innovative Behavior and Categories tests are not reliable. Performance measures resulting from these two instruments are not precise and conclusions based on these test scores should be considered tenuous.

The reliability of the Metropolitan Readiness Test total score is adequate. At three different testing sessions on 167, 173 and 200 children, the corrected odd-even reliability coefficients for the total test were .91, .91 and .94 respectively. Reliabilites of subtests were lower and the authors cautioned the user when working with subtest scores.

*Throughout this paper, "poor" children are children who meet the Head Start Guidelines for poverty as outlined by the Federal Government.

IV. Kindergarten Test Results by Community

This section gives a short report on the average test scores made by poor and non-poor kindergarten children in seven of eight participating communities. Communities are identified by letter and not by name. The test results for Community C's kindergarten children are not included in this report since there were testing difficulties and the accuracy of the test scores is questionable. The difficulties in testing are reflected in the childrens' changes on the WPPSI. Thirty-five of Community C's 51 poor Follow Through kindergarten children made lower WPPSI scores at the end of the year than they had at the beginning. Some decreases were as much as 16 points.

Also, it should be remembered that reported PSI scores are median percentiles. Percentile scores were calculated for both poor and non-poor children using middle-class norms.

Community A

Table 2 shows the test performance made by children in Community A.

As shown, both poor and non-poor children made considerable gains on WPPSI test. On the average, thirty-six poor children scored about six points below the national average intelligence of 40 at the beginning of the year. At the end of the year 33 of the children who were posttested made an Average WPPSI score of 40; an average six-point increase. Verbal gains were slightly higher than performance gains.

Although non-poor children in Community A also made gains on the WPPSI, they began the year testing far above the general population. The gains made over the six-month period place them approximately one standard deviation (10 pts.) above the intelligence level of children their age. Performance

TABLE 2

AVERAGE PRE, POST AND CHANGES ON THE WPPSI, PSI, IB AND C TEST MADE BY POOR AND NON-POOR FOLLOW THROUGH CHILDREN IN COMMUNITY A

GROUP	TEST	N	AVG. PRE SCORE	N	AVG. POST SCORE	CHANGE
POOR	WPPSI VERBAL	(36)	16.69	(33)	19.79	3.10
	WPPSI PERFORMANCE	(36)	17.28	(33)	19.70	2.42
	TOTAL WPPSI	(36)	33.97	(33)	39.79	5.82
	PRESCHOOL INV.*	(59)	5.00	(59)	15.00	10.00
	INN. BEH.	(24)	7.79	(21)	7.95	0.16
	C TEST	(24)	3.50	(21)	5.52	2.02
NON-POOR	WPPSI VERBAL	(37)	24.14	(32)	25.09	0.95
	WPPSI PERFORMANCE	(37)	23.32	(32)	26.25	2.93
	TOTAL WPPSI	(37)	47.46	(32)	50.72	3.26
	PRESCHOOL INV.*	(71)	95.00	(71)	95.00	0.00
	INN. BEH.	(29)	13.14	(25)	12.04	-1.10
	C TEST	(29)	5.62	(25)	7.20	1.58

*Represents median percentile scores.

on the PSI, reflected by medain percentile scores, increased for poor children. Since the non-poor children recorded such a high median percentile score on the pretest, changes for this group were negligible.

In all cases poor children tested lower at the beginning of the year than non-poor children. Note that the non-poor children decreased in Innovative Behavior score.

Community B

Test results for Community B are found in Table 3.

TABLE 3

AVERAGE PRE, POST AND CHANGES ON THE WPPSI, PSI, IB AND C TEST MADE BY POOR AND NON-POOR FOLLOW THROUGH CHILDREN IN COMMUNITY B

GROUP	TEST	N	AVG. PRE SCORE	N	AVG POST SCORE	CHANGE
POOR	WPPSI VERBAL	(90)	17.01	(79)	19.75	2.74
	WPPSI PERFORMANCE	(90)	16.24	(79)	19.19	2.95
	TOTAL WPPSI	(90)	33.26	(79)	39.04	5.78
	PRESCHOOL INV.	(74)	10.00	(74)	32.50	22.50
	INN. BEH.	(91)	7.39	(80)	10.88	3.49
	C TEST	(91)	4.42	(80)	6.16	1.74
NON-POOR	WPPSI VERBAL	(9)	14.78	(8)	18.75	3.97
	WPPSI PERFORMANCE	(9)	11.89	(8)	16.88	4.99
	TOTAL WPPSI	(9)	26.67	(8)	35.63	8.96
	PRESCHOOL INV	(6)	2.50	(6)	20.00	17.50
	INN. BEH.	(10)	7.10	(8)	7.88	0.78
	C TEST	(10)	3.40	(8)	5.25	1.85

As shown in Table 3 the WPPSI gains made by children in Community B were substantial. When pretested, the 90 poor children scored about 33 points. Posttest poor children increased by an average of six points.

WPPSI gains for non-poor children look markedly higher; however, they only represent scores made by nine children on the pretest and eight children on the posttest. With so few children the average scores could be influenced by one or two extremely high scores.

Preschool Inventory, Innovative Behavior and Categories test scores also increased for both groups. Over the year PSI group scores rose from the 10th to the 32nd percentile for the poor children and from the 2nd to the 20th percentile for the few non-poor children. A puzzling thing about Community B is that non-poor children were uniformly lower than poor children.

Community D

Test results for Community D are found in Table 4.

TABLE 4
AVERAGE PRE, POST AND CHANGES ON THE WPPSI, PSI, IB AND C TEST MADE BY POOR AND NON-POOR FOLLOW THROUGH CHILDREN IN COMMUNITY D

GROUP	TEST	N	AVG. PRE SCORE	N	AVG POST SCORE	CHANGE
POOR	WPPSI VERBAL	(31)	16.19	(31)	17.71	1.52
	WPPSI PERFORMANCE	(31)	18.55	(31)	21.23	2.68
	TOTAL WPPSI	(31)	34.71	(31)	38.94	4.23
	PRESCHOOL INV.	(30)	27.50	(30)	90.25	62.75
	INN. BEH.	(32)	7.94	(30)	12.27	4.33
	C TEST	(32)	4.91	(31)	6.16	1.25
NON-POOR	WPPSI VERBAL	(10)	15.50	(9)	16.89	1.39
	WPPSI PERFORMANCE	(10)	17.10	(9)	20.78	3.68
	TOTAL WPPSI	(10)	32.60	(9)	37.67	5.07
	PRESCHOOL INV	(9)	22.50	(9)	91.25	68.75
	INN. BEH.	(13)	10.77	(13)	15.15	4.38
	C TEST	(13)	5.15	(13)	6.85	1.70

Again, the overall patterns of increases from pretest to posttest is evident. Average increases in WPPSI scores are 4 points for the poor children and 5 points for the non-poor children. Both groups score about the same on the measure of intelligence used in this study.

Considerable gains are also made on the Innovative Behavior test as well as on the Preschool Inventory. Poor and non-poor children pretested around the 25th percentile on the PSI. At the end of the year, the increase in test performance placed them above the 90th percentile of middle-class children their age.

Community E

Table 5 shows the average pre and postest scores for this community.

TABLE 5

AVERAGE PRE, POST AND CHANGES ON THE WPPSI, PSI, IB AND C TEST MADE BY POOR AND NON-POOR FOLLOW THROUGH CHILDREN IN COMMUNITY E

GROUP	TEST	N	AVG. PRE SCORE	N	AVG POST SCORE	CHANGE
POOR	WPPSI VERBAL	(57)	14.14	(48)	20.83	6.69
	WPPSI PERFORMANCE	(57)	16.83	(48)	21.04	4.21
	TOTAL WPPSI	(57)	30.97	(48)	42.77	11.80
	PRESCHOOL INV.	(51)	11.65	(51)	23.75	12.10
	INN. BEH.	(57)	6.35	(49)	8.02	1.67
	C TEST	(60)	4.90	(52)	6.00	1.10
NON-POOR	WPPSI VERBAL	(4)	8.00	(4)	19.25	11.25
	WPPSI PERFORMANCE	(4)	15.00	(4)	22.75	7.75
	TOTAL WPPSI	(3)	25.33	(4)	42.00	16.67
	PRESCHOOL INV	(5)	3.33	(5)	15.00	11.67
	INN. BEH.	(4)	5.50	(4)	6.00	0.50
	C TEST	(4)	3.75	(4)	6.25	2.50

As shown, when 57 poor Follow Through children in this Community were pretested on intelligence, their average score was 31, about one standard deviation below the national average. Upon posttesting the average score of

48 of these children was 43, slightly above the national average for intelligence using this test. This difference represented a change of 12 points and reflects more of an increase in the verbal area than it does with items on the performance subtest.

Although non-poor children recorded an average WPPSI gain of 17 points, the gain reflects only three children on the pretest and four on the posttest.

Innovative Behavior percentile scores made at the end of the year by 49 poor children were high and reflected a gain of two points from the beginning of the year. Gains on the PSI are relatively modest.

Community F

Test results for children in Community F are found in Table 6.

TABLE 6

AVERAGE PRE, POST AND CHANGES ON THE WPPSI, PSI, IB AND C TEST MADE BY POOR AND NON-POOR FOLLOW THROUGH CHILDREN IN COMMUNITY F

GROUP	TEST	N	AVG. PRE SCORE	N	AVG POST SCORE	CHANGE
POOR	WPPSI VERBAL	(20)	16.95	(18)	18.67	1.72
	WPPSI PERFORMANCE	(20)	17.40	(18)	19.06	1.66
	TOTAL WPPSI	(20)	34.35	(18)	36.39	2.04
	PRESCHOOL INV.	(11)	14.25	(11)	22,50	8.25
	INN. BEH.	(20)	7.80	(17)	7.94	0.14
	C TEST	(20)	3.90	(17)	5.35	1.45
NON-POOR	WPPSI VERBAL	(21)	19.95	(20)	18.90	-1.05
	WPPSI PERFORMANCE	(21)	19.05	(20)	20.20	1.15
	TOTAL WPPSI	(21)	39.00	(20)	38.10	-0.90
	PRESCHOOL INV	(15)	55.00	(15)	82.50	27.50
	INN. BEH.	(23)	7.83	(21)	8.57	0.74
	C TEST	(23)	4.00	(21)	5.71	1.71

As Table 6 shows, the poor Follow Through children in Community F increased their average scores on all measures. The WPPSI scores increased on the average about two points and resulted in an average posttest score of 36 for the 18 children tested. A score of 36 is below the average for the general population of children.

The non-poor children scored about the same on the two administrations of the WPPSI. The average total WPPSI score for the 20 non-poor children tested at the end of the year is about one point lower than the average score for these same children at the beginning of the year. The median PSI score for the poor and non-poor children increased over the year.

With both groups of children, average scores made on the Categories test increased approximately two points from the beginning to the end of the year while the Innovative Behavior Test recorded negligible gains.

Community G

Results for Community G are found in Table 7.

TABLE 7
AVERAGE PRE, POST AND CHANGES ON THE WPPSI, PSI, IB AND C TEST MADE BY POOR AND NON-POOR FOLLOW THROUGH CHILDREN IN COMMUNITY G

GROUP	TEST	N	AVG. PRE SCORE	N	AVG POST SCORE	CHANGE
POOR	WPPSI VERBAL	(69)	18.00	(63)	19.92	1.92
	WPPSI PERFORMANCE	(69)	17.68	(63)	20.27	2.59
	TOTAL WPPSI	(69)	35.54	(63)	40.19	4.65
	PRESCHOOL INV.	(55)	50.00	(55)	65.00	15.00
	INN. BEH.	(76)	5.90	(65)	7.71	1.81
	C TEST	(76)	4.61	(62)	5.73	1.12
NON-POOR	WPPSI VERBAL	(33)	17.91	(31)	20.77	2.86
	WPPSI PERFORMANCE	(33)	19.36	(31)	21.84	2.48
	TOTAL WPPSI	(32)	36.66	(31)	42.61	5.95
	PRESCHOOL INV	(32)	72.50	(32)	85.00	12.50
	INN. BEH.	(35)	5.71	(33)	7.09	1.38
	C TEST	(35)	5.26	(33)	5.70	0.44

Intelligence scores in Community G were an average five points higher at the end of the school year for approximately 100 poor and non-poor children. This increase resulted in an average posttest WPPSI score of 43 for the non-poor children and 40 for the poor children. PSI median scores also increased about equally for both groups as did scores made on the Innovative Behavior subtest.

Both poor and non-poor children made slight gains on the Categories Test.

Community H

There were no non-poor Follow Through kindergarten children in Community H. Test results for poor kindergarten children at this site are found in Table 8.

TABLE 8

AVERAGE PRE, POST AND CHANGES ON THE WPPSI, PSI, IB AND C TEST MADE BY POOR FOLLOW THROUGH CHILDREN IN COMMUNITY H

GROUP	TEST	N	AVG. PRE SCORE	N	AVG POST SCORE	CHANGE
POOR	WPPSI VERBAL	(24)	22.00	(26)	23.42	1.42
	WPPSI PERFORMANCE	(24)	16.75	(26)	17.81	1.06
	TOTAL WPPSI	(24)	38.75	(26)	41.23	2.48
	PRESCHOOL INV.	(16)	11.00	(16)	24.00	13.00
	INN. BEH.	(29)	5.07	(23)	9.30	4.23
	C TEST	(30)	3.77	(23)	7.39	3.62

Poor children in Community H also demonstrated positive changes on all measures. The WPPSI scores increased on the average approximately 3 points. The 26 children posttested on the WPPSI at the end of the year score slightly above the national average on this test.

Similar increases over the school year were made on the PSI. Innovative Behavior and Categories test scores showed large increases.

Summary of Kindergarten Test Results

Since the WPPSI is becoming one of the most reputable national instruments for measuring intellectual capacity in children aged four to six and a half, it seems appropriate to summarize how well poor and non-poor children in Responsive Environment classrooms performed on the four subtests of this instrument. Table 9 does this.

TABLE 9

A SUMMARY OF TOTAL WPPSI PRE- AND POSTTEST SCORES FOR POOR AND NON-POOR CHILDREN IN SEVEN COMMUNITIES

Community	Group	(N)	Pre Score*	(N)	Post Score*	Change
A	poor	(36)	34	(33)	40	6
	non-poor	(37)	47	(32)	51	4
B	poor	(90)	33	(79)	39	6
	non-poor	(9)	27	(8)	36	9
D	poor	(31)	35	(31)	39	4
	non-poor	(10)	33	(9)	38	5
E	poor	(57)	31	(48)	43	12
	non-poor	(3)	25	(4)	42	17
F	poor	(20)	34	(18)	36	2
	non-poor	(21)	39	(20)	38	-1
G	poor	(69)	36	(63)	40	4
	non-poor	(32	37	(31)	43	6
H	poor	(24)	39	(26)	41	2
	non-poor	(non available)				
TOTAL (weighted)	poor	(327)	34	(298)	40	6
	non-poor	(112)	39	(104)	43	4

* Average pre and posttest scores were rounded to the nearest whole number.

As indicated in Table 9, average WPPSI intelligence scores made by poor groups of children increased in every community. The average increase for

poor children in all communities was six points (see Figure 1).

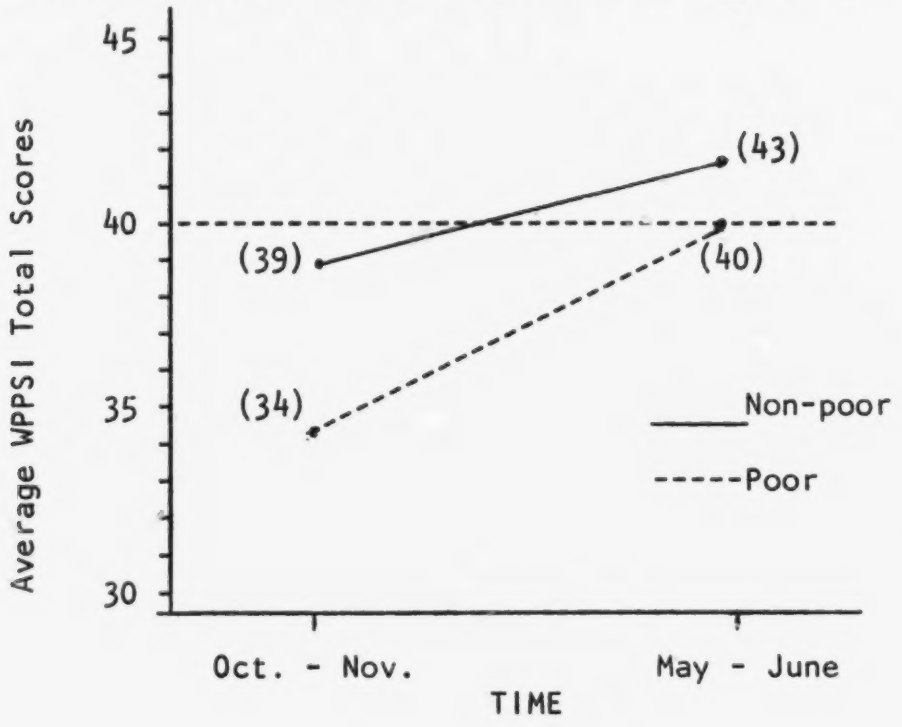

Figure 1. Changes between average pretest and average posttest WPPSI scores for poor and non-poor Follow Through children.

Poor children on the average scored 34 at the beginning of the year. This was slightly below the national average level of intelligence. At the end of the year-long kindergarten experience, their intelligence scores were equivalent to the national average of 40.

Non-poor children in the Responsive Environment classrooms made less change in WPPSI scores throughout the year. These children, however, began the school year at a higher level of intelligence as measured by the WPPSI. One hundred and twelve non-poor children made an average score of 39 on the WPPSI. The posttest average score of 104 of these children was 43.

There was considerable variability on WPPSI group changes made by poor and non-poor children in different communities. As shown in Table 9 and Figure 2, the poor and non-poor groups in Community E increased 12 and 17 WPPSI points respectively.

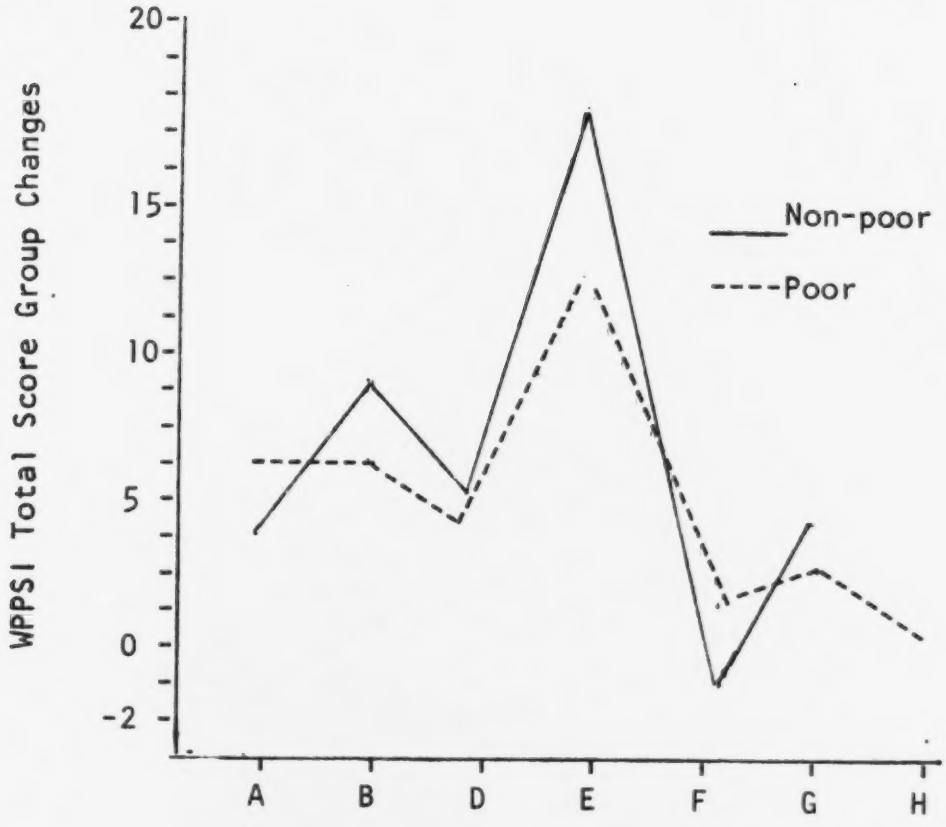

Figure 2. Changes between pretest and posttest poor and non-poor Follow Through groups in seven communities.

In Community B, where children pretested about the same as children in Community E group changes were on the average 6 points for the poor group and 9 points for the non-poor group. At only one site, Community F, there was a decrease in WPPSI scores from pretest to posttest. This decrease was one point for non-poor children.

Providing a general overall summary of performance by kindergarten children on the Preschool Inventory is difficult. A kindergarten child's score on the PSI was transferred to a percentile score using middle-class norms for both poor and non-poor children. Median percentile scores were then reported for poor and non-poor groups in each community. Because test protocols were returned to their respective communities, raw PSI scores could not be retrieved and analyzed to provide a more comprehensive summary.

However, Table 10 shows the median PSI percentile score for poor and non-poor children in the kindergarten Follow Through communities.

TABLE 10

MEDIAN PRESCHOOL INVENTORY PRE- AND POSTTEST SCORES AND CHANGES FOR POOR AND NON-POOR CHILDREN IN SEVEN COMMUNITIES USING MIDDLE CLASS NORMS

COMMUNITY	GROUP	N	PRETEST MDN. SCORE	POSTTEST MDN. SCORE	MEDIAN INCREASE
A	Poor	59	5.00	15.00	+10.00
	Non-Poor	71	95.00	95.00	0.00
B	Poor	74	10.50	32.50	+22.50
	Non-Poor	6	2.50	20.00	+17.50
D	Poor	30	27.50	90.25	+62.75
	Non-Poor	9	22.50	91.25	+68.75
E	Poor	51	11.65	23.75	+12.10
	Non-Poor	5	3.33	15.00	+11.67
F	Poor	11	14.25	22.50	+ 8.25
	Non-Poor	15	55.00	82.50	+27.50
G	Poor	55	50.00	65.00	+15.00
	Non-Poor	32	72.50	85.00	+12.50
H	Poor	16	11.00	24.00	+13.00
	Non-Poor	*			
TOTAL	Poor	296	11.80	26.10	+14.30
	Non-Poor	138	22.50	72.50	+50.00

*There were no non-poor Follow Through children in Community H.

As shown, median scores increased approximately 14 percentile points for 296 poor children and 50 points for the 138 non-poor children. When analyzed by communities, it will be noticed that some groups, such as the poor and non-poor in Community D, showed considerable increase in median percentile scores. Another group, the non-poor in Community A, because the scores were so high on the pretest, showed no median increase.

In Communities B and E the posttest PSI median scores was higher for the poor group. No conclusion can be made however, because of the few children in the non-poor groups in these two communities.

V. First Grade Test Results

Two communities had first grade Follow Through classes in 1968. As was discussed, the Metropolitan Readiness Test was administered to all children in each of these communities at the beginning and end of the year Test scores were not reported separately for economic levels, since almost all children in Communities I and J are poor. Table 11 shows the average Metropolitan raw scores for first graders in these communities.

TABLE 11
FIRST GRADE METROPOLITAN PRE-AND POSTTEST RAW SCORES

		Average Scores		
Communities	N	Pretest	Posttest	Change
I	293	39.63	66.10	26.57
J	124	41.15	71.96	30.81
Total(weighted)	417	40.88	67.84	27.76

As indicated above, first grade children in each community performed relatively low on the pretest. The children scored on the average about 40 out of the 102 total possible score. A score of 40 corresponds to the 23rd percentile rank for children in the standarization sample. According ot the test developers a score of 40 would categorize a child as having a "Low Normal readiness status". That is, that child is "likely to have difficulty in first grade work" and "should be assingned to a slow section and given more individualized help." This low score is understandable knowing 85% of the children in Communities I and J are children from disadvantaged homes.

At the end of their first grade experience the children scored considerably higher on the Metropolitan. The average posttest raw score for the 417 first graders tested was 68. A score of 68 falls at the 77th percentile on norms for beginning first grade children. However, considering the low socio-economic status of the children in Community I and J, and the fact that the Metropolitan norming group was somewhat biased toward higher income families, it is reasonable to conclude that first graders in Communities I and J would compare very favorably to norms calculated at the end of the first grade experience if they were available.

A child with a raw score of 68 would be classified by the test developers (at the beginning of the first grade experience) as having "high normal readiness status." This child would be characterized as having "good prospects for success in first-grade work provided other indications, such as health, emotional factors, etc. are consistent."

VI. Overview

This paper is a preliminary report of objective test results made by approximately 300 poor and 100 non-poor kindergarten children and by 417 first grade children during 1968-1969. The test results were used to measure changes in a child's intellectual ability as a result of one year's schooling in a Responsive Follow Through classroom. Except for four instances, average test scores made by groups of both poor and non-poor kindergarten Follow Through children increased for all tests in every community. That is, over the seven to nine month time period, kindergarten children in all communities increased in intelligence, the ability to form concepts, the ability to solve problems and the ability to categorize. Results for first grade children were also encouraging. There was considerable increases in Metropolitan test scores for both first grade communities from the beginning to end of the school year.

The test results reported in this paper were summarized by grade within the community. For each community of kindergarten children, the performance of children classified as poor was separate from performance made by non-poor children. These breakdowns reflect the least number of ways the test results had to be categorized before they could be meaningful; but this is a limitation.

There are other ways to summarize children's scores. For example, a more important question could be answered by determining children's performance as related to their year-long experiences in a specific Responsive classroom. To answer this question, the degree to which the Responsive model has been carried out in each classroom would first be determined. Since we already know that in different communities poor and non-poor children score differently on the tests, the analysis would then take into consideration variable of grade, community, poor vs. non-poor, and the degree to which the model has been implemented by a teacher and her assistant.

Using the results obtained by the Classroom Observation Schedule, an index of the degree of "responsiveness" will be figured for each classroom. This variable will then be correlated with test performance. Other analyses must also be made in examining the effectiveness of the Responsive model. For example, questions must be answered dealing with the model's effectiveness with different ethnic groups, for different sexes, or for children where the primary language in the home is not English. Biographic and demographic information provided on the Child Data Sheet will help us answer these questions.

Undoubtedly the major questions that must eventually be resolved are concerned with the long-term effects of the program and how children in the Responsive Follow Through Model compare with children in other Follow Through models or in conventional classroom settings. These questions must take into account attitude toward school, self-concept related to school, along with concept formation, and problem-solving abilities. Answers to these questions will influence not only the nature of early childhood education, but the nature of all education. For this reason, evaluation in the Responsive Model must continue to be an integral part of the program.

Nick Rayder
Ann Fitzgibbon
Margery Sakai
Randy Hake
Keith Alward
Debora Walker

3/8/70